TRhizo-urbanMicrobiome
Mutualistic Bacteria

David Murray-Stoker

Contents

Load Packages

Load Data

```r
## Microbiome tidyamplicons
microbiome.tidyamplicon.reference <- read_rds(
  file = "data/microbiome_tidyamplicon_reference.rds"
)

## Load the urbanization data
urbanization.data <- read_rds(
  file = "data/urbanization_data.rds"
)

## Load the workspace
load("data_analysis/10-bacterial_functional_groups/mutualistic_bacteria-workspace.RData")
```

Data Management

```r
## Set reference ASV abundance matrix
ASV.abundance.matrix <- add_rel_abundance(microbiome.tidyamplicon.reference) %>%
  abundances()

## Set reference taxa matrix
ASV.taxa.matrix <- taxa(microbiome.tidyamplicon.reference)

## Set tibble of bacteria abundance
bacteria.abundance.data <- ASV.abundance.matrix %>%
  full_join(ASV.taxa.matrix, by = "taxon_id") %>%
  full_join(microbiome.tidyamplicon.reference$samples, by = "sample_id") %>%
  select(Population:Compartment, phylum:genus, abundance, rel_abundance) %>%
  rename(
    Phylum = phylum, Class = class, Order = order, Family = family, Genus = genus,
    Abundance = abundance, Relative_Abundance = rel_abundance
  )

## Set aggregated sample data
sample.data <- microbiome.tidyamplicon.reference$samples %>%
  select(Population:Compartment) %>%
  distinct()
```

Filter Abundances

```r
## Set tibbles of abundances for different mutualistic bacteria
# Rhizobium
rhizobium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Rhizobium") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Rhizobium_Abundance = sum(Abundance),
    Rhizobium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Allorhizobium
allorhizobium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Allorhizobium") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Allorhizobium_Abundance = sum(Abundance),
    Allorhizobium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Bradyrhizobium
bradyrhizobium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Bradyrhizobium") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Bradyrhizobium_Abundance = sum(Abundance),
    Bradyrhizobium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Mesorhizobium
mesorhizobium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Mesorhizobium") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Mesorhizobium_Abundance = sum(Abundance),
    Mesorhizobium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Pararhizobium
pararhizobium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Pararhizobium") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
```

```r
    Pararhizobium_Abundance = sum(Abundance),
    Pararhizobium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Shinella
shinella.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Shinella") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Shinella_Abundance = sum(Abundance),
    Shinella_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


## Combine abundances into a tibble with urbanization metrics
mutualistic.bacteria.data <- sample.data %>%
  full_join(rhizobium.abundances, by = c("Population", "Compartment")) %>%
  full_join(allorhizobium.abundances, by = c("Population", "Compartment")) %>%
  full_join(bradyrhizobium.abundances, by = c("Population", "Compartment")) %>%
  full_join(mesorhizobium.abundances, by = c("Population", "Compartment")) %>%
  full_join(pararhizobium.abundances, by = c("Population", "Compartment")) %>%
  full_join(shinella.abundances, by = c("Population", "Compartment")) %>%
  replace(is.na(.), 0) %>%
  full_join(urbanization.data, by = "Population") %>%
  type_convert(col_types = c("ffnnnnnnnnnnnnnnnn"))
```

Rhizobium GAMs

```r
## Rhizobium by distance
rhizobium.by.distance.GAM <- gam(
  Rhizobium_Relative_Abundance ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Rhizobium by HII
rhizobium.by.HII.GAM <- gam(
  Rhizobium_Relative_Abundance ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Rhizobium by ISC
rhizobium.by.ISC.GAM <- gam(
  Rhizobium_Relative_Abundance ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)
```

Check Model Assumptions

```r
## Rhizobium abundance-by-distance model diagnostics
check_model(rhizobium.by.distance.GAM)
# Visual check = assumptions met

## Rhizobium abundance-by-HII model diagnostics
check_model(rhizobium.by.HII.GAM)
# Visual check = assumptions met

## Rhizobium abundance-by-ISC model diagnostics
check_model(rhizobium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 1: ANOVA table for the Rhizobium relative abundance-by-distance GAM. Adjusted R-squared = 0.646, deviance = 66.8 Compartment: F = 109.3, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	3.196	9	1.956	0.001
s(Distance):CompartmentSoil	0.000	9	0.000	0.995

Table 2: ANOVA table for the Rhizobium relative abundance-by-HII GAM. Adjusted R-squared = 0.564, deviance = 57.4 Compartment: F = 88.7, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.598	9	0.165	0.119
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.967

Table 3: ANOVA table for the Rhizobium relative abundance-by-ISC GAM. Adjusted R-squared = 0.734, deviance = 76.1 Compartment: F = 145.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	5.959	9	5.115	0.000
s(Mean_ISC):CompartmentSoil	0.000	9	0.000	0.997

Allorhizobium GAMs

```r
## Allorhizobium by distance
allorhizobium.by.distance.GAM <- gam(
  Allorhizobium_Relative_Abundance ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)


## Allorhizobium by HII
allorhizobium.by.HII.GAM <- gam(
  Allorhizobium_Relative_Abundance ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)


## Allorhizobium by ISC
allorhizobium.by.ISC.GAM <- gam(
  Allorhizobium_Relative_Abundance ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)
```

Check Model Assumptions

```r
## Allorhizobium abundance-by-distance model diagnostics
check_model(allorhizobium.by.distance.GAM)
# Visual check = assumptions met

## Allorhizobium abundance-by-HII model diagnostics
check_model(allorhizobium.by.HII.GAM)
# Visual check = assumptions met

## Allorhizobium abundance-by-ISC model diagnostics
check_model(allorhizobium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 4: ANOVA table for the Allorhizobium relative abundance-by-distance GAM. Adjusted R-squared = 0.333, deviance = 36.9 Compartment: F = 12.7, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	2.644	9	2.535	0
s(Distance):CompartmentSoil	0.000	9	0.000	1

Table 5: ANOVA table for the Allorhizobium relative abundance-by-HII GAM. Adjusted R-squared = 0.218, deviance = 24.0 Compartment: F = 10.8, P = 0.002.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.905	9	1.051	0.002
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.997

Table 6: ANOVA table for the Allorhizobium relative abundance-by-ISC GAM. Adjusted R-squared = 0.134, deviance = 15.5 Compartment: F = 9.781, P = 0.003.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.671	9	0.215	0.093
s(Mean_ISC):CompartmentSoil	0.000	9	0.000	1.000

Bradyrhizobium GAMs

```r
## Bradyrhizobium by distance
bradyrhizobium.by.distance.GAM <- gam(
  Bradyrhizobium_Relative_Abundance ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Bradyrhizobium by HII
bradyrhizobium.by.HII.GAM <- gam(
  Bradyrhizobium_Relative_Abundance ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Bradyrhizobium by ISC
bradyrhizobium.by.ISC.GAM <- gam(
  Bradyrhizobium_Relative_Abundance ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)
```

Check Model Assumptions

```r
## Bradyrhizobium abundance-by-distance model diagnostics
check_model(bradyrhizobium.by.distance.GAM)
# Visual check = assumptions met

## Bradyrhizobium abundance-by-HII model diagnostics
check_model(bradyrhizobium.by.HII.GAM)
# Visual check = assumptions met

## Bradyrhizobium abundance-by-ISC model diagnostics
check_model(bradyrhizobium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 7: ANOVA table for the Bradyrhizobium relative abundance-by-distance GAM. Adjusted R-squared = 0.371, deviance = 39.7 Compartment: F = 36.4, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.000	9	0.000	0.720
s(Distance):CompartmentSoil	1.894	9	0.589	0.055

Table 8: ANOVA table for the Bradyrhizobium relative abundance-by-HII GAM. Adjusted R-squared = 0.322, deviance = 33.2 Compartment: F = 33.8, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0	9	0	0.675
s(Human_Influence_Index):CompartmentSoil	0	9	0	0.413

Table 9: ANOVA table for the Bradyrhizobium relative abundance-by-ISC GAM. Adjusted R-squared = 0.374, deviance = 40.1 Compartment: F = 36.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.000	9	0.000	0.949
s(Mean_ISC):CompartmentSoil	1.948	9	0.634	0.048

Mesorhizobium GAMs

```r
## Mesorhizobium by distance
mesorhizobium.by.distance.GAM <- gam(
  Mesorhizobium_Relative_Abundance ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)


## Mesorhizobium by HII
mesorhizobium.by.HII.GAM <- gam(
  Mesorhizobium_Relative_Abundance ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)


## Mesorhizobium by ISC
mesorhizobium.by.ISC.GAM <- gam(
  Mesorhizobium_Relative_Abundance ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)
```

Check Model Assumptions

```r
## Mesorhizobium abundance-by-distance model diagnostics
check_model(mesorhizobium.by.distance.GAM)
# Visual check = assumptions met

## Mesorhizobium abundance-by-HII model diagnostics
check_model(mesorhizobium.by.HII.GAM)
# Visual check = assumptions met

## Mesorhizobium abundance-by-ISC model diagnostics
check_model(mesorhizobium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 10: ANOVA table for the Mesorhizobium relative abundance-by-distance GAM. Adjusted R-squared = 0.320, deviance = 33.2 Compartment: F = 33.1, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.278	9	0.043	0.244
s(Distance):CompartmentSoil	0.000	9	0.000	0.443

Table 11: ANOVA table for the Mesorhizobium relative abundance-by-HII GAM. Adjusted R-squared = 0.392, deviance = 42.1 Compartment: F = 37.0, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.046	9	0.005	0.311
s(Human_Influence_Index):CompartmentSoil	2.229	9	0.936	0.016

Table 12: ANOVA table for the Mesorhizobium relative abundance-by-ISC GAM. Adjusted R-squared = 0.326, deviance = 34.1 Compartment: F = 33.4, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.5	9	0.11	0.163
s(Mean_ISC):CompartmentSoil	0.0	9	0.00	0.713

Pararhizobium GAMs

```r
## Pararhizobium by distance
pararhizobium.by.distance.GAM <- gam(
  Pararhizobium_Relative_Abundance ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Pararhizobium by HII
pararhizobium.by.HII.GAM <- gam(
  Pararhizobium_Relative_Abundance ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Pararhizobium by ISC
pararhizobium.by.ISC.GAM <- gam(
  Pararhizobium_Relative_Abundance ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)
```

Check Model Assumptions

```r
## Pararhizobium abundance-by-distance model diagnostics
check_model(pararhizobium.by.distance.GAM)
# Visual check = assumptions met

## Pararhizobium abundance-by-HII model diagnostics
check_model(pararhizobium.by.HII.GAM)
# Visual check = assumptions met

## Pararhizobium abundance-by-ISC model diagnostics
check_model(pararhizobium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 13: ANOVA table for the Pararhizobium relative abundance-by-distance GAM. Adjusted R-squared = 0.675, deviance = 68.3 Compartment: F = 142.8, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.612	9	0.172	0.116
s(Distance):CompartmentSoil	0.000	9	0.000	0.785

Table 14: ANOVA table for the Pararhizobium relative abundance-by-HII GAM. Adjusted R-squared = 0.671, deviance = 67.8 Compartment: F = 141.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.424	9	0.082	0.192
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.896

Table 15: ANOVA table for the Pararhizobium relative abundance-by-ISC GAM. Adjusted R-squared = 0.670, deviance = 67.7 Compartment: F = 140.8, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.357	9	0.061	0.219
s(Mean_ISC):CompartmentSoil	0.000	9	0.000	0.853

Shinella GAMs

```r
## Shinella by distance
shinella.by.distance.GAM <- gam(
  Shinella_Relative_Abundance ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Shinella by HII
shinella.by.HII.GAM <- gam(
  Shinella_Relative_Abundance ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)

## Shinella by ISC
shinella.by.ISC.GAM <- gam(
  Shinella_Relative_Abundance ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = mutualistic.bacteria.data,
  method = "REML"
)
```

Check Model Assumptions

```r
## Shinella abundance-by-distance model diagnostics
check_model(shinella.by.distance.GAM)
# Visual check = assumptions met

## Shinella abundance-by-HII model diagnostics
check_model(shinella.by.HII.GAM)
# Visual check = assumptions met

## Shinella abundance-by-ISC model diagnostics
check_model(shinella.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 16: ANOVA table for the Shinella relative abundance-by-distance GAM. Adjusted R-squared = 0.391, deviance = 40.8 Compartment: F = 40.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.869	9	0.579	0.016
s(Distance):CompartmentSoil	0.000	9	0.000	0.997

Table 17: ANOVA table for the Shinella relative abundance-by-HII GAM. Adjusted R-squared = 0.431, deviance = 44.6 Compartment: F = 42.9, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.911	9	1.138	0.001
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	1.000

Table 18: ANOVA table for the Shinella relative abundance-by-ISC GAM. Adjusted R-squared = 0.366, deviance = 38.2 Compartment: F = 38.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.708	9	0.25	0.078
s(Mean_ISC):CompartmentSoil	0.000	9	0.00	0.996

Export Data

```r
## Mutualistic bacteria
write_rds(
  mutualistic.bacteria.data,
  file = "data/mutualistic_bacteria_data.rds"
)
```

```r
save.image("data_analysis/10-bacterial_functional_groups/mutualistic_bacteria-workspace.RData")
```

R Session Information

Table 19: Packages required for data management and analysis.

Package	Loaded Version	Date
bayestestR	0.13.2	2024-02-12
broom	1.0.5	2023-06-09
correlation	0.8.4	2023-04-06
datawizard	0.10.0	2024-03-26
dplyr	1.1.4	2023-11-17
easystats	0.7.1	2024-03-28
effectsize	0.8.7	2024-04-01
forcats	1.0.0	2023-01-29
ggplot2	3.5.1	2024-04-23
insight	0.19.10	2024-03-22
kableExtra	1.4.0	2024-01-24
knitr	1.46	2024-04-06
lubridate	1.9.3	2023-09-27
mgcv	1.9-1	2023-12-21
modelbased	0.8.7	2024-02-15
nlme	3.1-164	2023-11-27
parameters	0.21.6	2024-03-18
performance	0.11.0	2024-03-22
purrr	1.0.2	2023-08-10
readr	2.1.5	2024-01-10
report	0.5.8	2023-12-07
see	0.8.4	2024-04-29
stringr	1.5.1	2023-11-14
tibble	3.2.1	2023-03-20
tidyamplicons	0.2.2	2022-09-10
tidyr	1.3.1	2024-01-24
tidyverse	2.0.0	2023-02-22